Research Article

THE RETRIEVABILITY OF L2 ENGLISH MULTIWORD ITEMS IN A CONTEXT OF STRONGLY FORM-FOCUSED EXPOSURE

WHAT MATTERS?

Seth Lindstromberg* ⓘD

Hilderstone College, United Kingdom

June Eyckmans

Ghent University, Belgium

Although retrieval of lexical forms is a prerequisite for language production, research of L2 vocabulary learning has focused much more on meanings and form-meaning mappings than on development of detailed, accessible mental representations of forms. This is particularly true with respect to multi-word items (MWIs). We report an experimental study involving a variety of intra-lexical, usage-based, and interlingual co-determinants of L2 vocabulary learnability pertaining to MWIs. Each learner (N = 60) encountered a randomly allocated set of 26 two-word MWIs (Nsets = 4) semi-randomly drawn from a larger pool of MWIs. Learners were asked to remember either the 13 MWIs showing the form variable assonance (e.g., *change shape*) or the 13 nonassonant control MWIs (e.g., *sound good*). Posttests of form recall revealed a large, durable effect of the focusing task in combination with forewarning of testing. Except when MWI concreteness (a semantic variable) was high, assonance had a positive effect on retrievability in recall tests given after delays of 15 minutes and one week. There was a consistent effect of the semi-semantic variable Mutual Information. Even in the context of a strong focus on forms, form variables are not the only variables that matter.

---

* Correspondence concerning this article should be addressed to Seth Lindstromberg, Hilderstone College, St Peters Road, Broadstairs, Kent, CT10 2JW, United Kingdom. E-mail: lindstromberg@gmail.com

Although language production presupposes an ability to retrieve the forms of L2 words whose meanings have been learned, form-learning has received a good deal less attention in research of second language vocabulary acquisition (SLVA) than meaning-learning (Barcroft, 2015, p. 115; Schmitt, 2008, pp. 336–337). Issues relating to the development of an ability to retrieve the forms of L2 multiword items (MWIs) may be in particular need of research attention given the evidence that development of productive knowledge of L2 MWIs tends to proceed slowly in comparison to productive knowledge of L2 words (Boers, 2020; Durrant & Schmitt, 2009; Forsberg, 2010; Laufer & Waldman, 2011; Li & Schmitt, 2010; Nekrasova, 2009; Nesselhauf, 2003; Yamashita & Jiang, 2010). One proposed approach to helping learners develop detailed, accessible mental representations of MWI forms is motivated in part by the fact that certain potentially mnemonic patterns of interword phonological similarity (aka sound repetition) are appreciably common in English MWIs of some types, especially in figurative idioms (e.g., *bite the bullet*, *buy time*) and sayings (*no pain no gain*) but also in comparatively literal and transparent collocations (*make a mess*, *state a case*) and in compound nouns (*force field*, *conspiracy theory*). One of these patterns is alliteration, as in *cold comfort* (Boers & Lindstromberg, 2009; Gries, 2011; Lindstromberg, 2020). Another is assonance, defined as the occurrence of a particular vowel phoneme (i.e., a monophthong, diphthong, or triphthong) in the most phonologically prominent syllable of at least two content words within a MWI, as in *bubble gum* (simple assonance), *a quick fix* (near or slant rhyme, where a postvocalic consonant is repeated), *fly high* (clipped rhyme), and *doom and gloom* (full rhyme). By this definition assonance occurs either on its own or as a component of the aforementioned types of rhyme. Importantly, the definition depends on the commonly made assumption that phonological similarity effects—in particular, effects on retrieval—may arise despite the presence of minor phonetic differences such as a difference in vowel length conditional on the length of the following consonant (e.g., Gupta et al., 2005). If such minor differences do matter, an effect presumed to operate at the level of the phoneme may turn out to be undetectable or else it may be so small as to be of little practical significance. Finally, because the present study involves only monosyllabic constituent words, it sidesteps the issue of comparative syllable prominence in di- or polysyllabic MWIs such as *high profile*, where the two syllables in *profile* might be judged to be similarly prominent.

An important point is whether assonance is common enough in English phraseology to be of practical significance for learners, teachers, and materials creators. We believe it is. For example, Boers et al. (2014a) read through *Oxford Idioms Dictionary for Learners of English* (Parkinson, 2006) tallying all defined idioms that include, in addition to any verbs, at least two content words of another class. Those researchers reported that of the 2,906 such expressions that they found, 392 (13.5%) show either simple assonance or a type of rhyme. Lindstromberg (2020) reported that of the 187 currently used binomials defined in Parkinson (2006) nearly 11% show simple assonance (*cut and run*) and a further 6% show a grade of rhyme (*high and dry*; *make or break*). For comparison, nearly 26% show alliteration. As just more than 3% of these binomials show assonance in combination with alliteration (*part and parcel of*), 39% manifest a pattern of sound repetition that may have a mnemonic effect. To give a final example, Boers et al. (2014a) examined 1,000 mostly comparatively literal Adj-N MWIs (*best friend*) formed from the

100 most frequent monosyllabic adjectives in the Corpus of Contemporary American English (COCA) (Davies, 2008 to present) and the 10 most frequent monosyllabic common noun collocates of each adjective.[1] Those researchers found (as slightly corrected by Lindstromberg, 2020) that 10% are assonant, including ones showing any type of rhyme (although ones that rhyme are rare). It may be of interest that in such expressions the by-chance incidence of any form of assonance is around 6%. The difference between the two proportions is statistically significant: $p < .001$ (Lindstromberg, 2020).

There is ample evidence that learners find it comparatively easy to recall *alliterative* MWIs (e.g., Boers & Lindstromberg, 2005, 2009; Boers et al., 2012; Boers et al., 2014c; Green, 2019), although the magnitude of a mnemonic effect of alliteration may depend on MWIs being made the object of form-focused attention direction (Boers et al., 2014b). For *assonance*, though, results have been mixed (Lindstromberg & Eyckmans, 2014, 2017). Moreover, it could be argued that past findings of a positive effect of assonance were artifacts of the attention direction tasks that were used: In some studies these tasks were preceded by awareness raising about assonance and, in most studies, the tasks were versions of sorting target MWIs according to whether they were thought to be assonant or not. Additionally, (a) no past investigation of a potential mnemonic effect of assonance controlled for more than two or three of the numerous lexical and semilexical variables that may also influence MWI memorability; (b) in all past studies the samples of assonant MWIs and nonassonant control MWIs were small (i.e., $n_{\text{assonant}} = n_{\text{control}} =$ 10 to 14); and (c) stimulus MWIs were never randomly selected from a pool of candidate MWIs.

For orientation through the next section of this article, a brief preview may be in order. To begin, the original core goal of our study was to address the following question: If assonance can make MWIs comparatively easy to learn and retrieve from memory, does it do so regardless of the focusing task, item concreteness, and item frequency? Over time the scope of the study was widened to address additional research questions that we come to in a later section. In relation to these questions a range of additional potential moderating variables were taken into account. Except for the focusing task, all the original and added independent variables are discussed in the following section. Because no published measures were available for three of the independent variables, we collected new measures in the form of subjective ratings. These we obtained either from known informants or from unknown informants using the crowd-sourcing platform Amazon Mechanical Turk as detailed in the following text. In the treatment, each learner (near immediate test: $N = 60$; 1-week delayed test: $N = 56$) engaged with one of four same-size randomly ordered lists of two-word stimulus MWIs ($N_{\text{total}} = 104$; $n_{\text{per.list}} = 26$). In each list, 13 of the MWIs assonate (e.g., *late stage*) and 13 show no seemingly mnemonic pattern of sound repetition. The lists, which were randomly allotted to learners, comprised MWIs that had been semirandomly drawn from larger pools of comparatively literal and transparent candidate MWIs deemed highly likely to be familiar to the participating learners, who were Dutch-speaking university undergraduates majoring in two foreign languages, one of which was English. Importantly, these MWIs were presented out of context; and the only context for each constituent word (CW) was the minimal context of the stimulus MWI it occurred in.

LEXICAL AND SEMILEXICAL VARIABLES FIGURING IN THE STUDY

VARIABLES OF LEXICAL FORM

Among the form-based variables figuring in our study are two "length" variables that have figured especially prominently in tasks used in experimental studies of memory for L1 lexical forms (e.g., lexical decision tasks, object and picture naming tasks, and tests of ability to recall or recognize previously studied paired associates). The first of these variables is orthographic length (i.e., number of letters); the second is phonological length (number of phonemes). Item length in *syllables* does not figure as a variable in the present study because, as mentioned, all stimulus MWIs consist of two monosyllabic words. Although it is fairly common for studies of SLVA to take account of length, findings have been mixed and the generality of length effects on vocabulary learnability remains unclear (Laufer, 1997; Peters, 2020). To measure the length of a MWI, we summed the lengths of its two CWs. The correlation between the orthographic and phonemic length of the MWIs in our study is $r = .37$.[2]

A second type of form variable figuring in many studies of L1 vocabulary acquisition and lexical processing but rarely in SLVA research is "neighborhood size" (NS), which we define shortly. We took account of *three* types of NS: English orthographic NS (OrthNS), English phononological NS (PhonNS), and Dutch Phon NS. OrthNS is the number of other words in a given language that differ from a specified word by the substitution, deletion, or addition of a single letter. For example, *sat* has the neighbors *mat*, *sit*, and *sad*. PhonNS is the number of other words that differ by one phoneme from the word of interest. For a MWI, our measure of NS was the sum of the NSs of its two CWs. OrthNS and PhonNS correlate positively with the extent to which individuals have distinct, detailed mental representations of lexical form (Storkel, 2004; for background see Marian et al., 2012; Yap & Balota, 2015). In a study involving three word-learning tasks Stamer and Vitevitch (2012) found that L2 Spanish words with high PhonNS were more learnable than ones with low PhonNS. We included *Dutch* PhonNS because L2 phonological forms that are similar to many L1 forms may present relatively few problems related to pronunciation. This may matter because pronunciability is positively associated with word learnability (e.g., Ellis & Beaton, 1993). For our MWIs the three measures of NS correlate as follows: OrthNS—PhonNS, $r = .56$; OrthNS—DutchNS, $r = .22$; PhonNS—DutchPhonNS, $r = .35$. To obtain measures of all the previously mentioned form variables we drew on the online multilingual Clearpond database (Marian et al., 2012).

A form variable of a third type, ±assonance, has already been introduced. We should add that a mnemonic effect of assonance might arise when retrieval of the form of a previously encountered word string (e.g., *stone cold*) is facilitated by "phonological priming," whereby recalling or remeeting one word in the string activates the memory traces of any other words in the string that are phonologically similar (for background Goldinger et al., 1992; Luce et al., 2000; Lupker & Williams, 1989). However, effects of phonological priming have most often been observed with respect to the onsets and the ends of words as in, respectively, alliteration and clipped or full rhyme. Thus, finding firm evidence of a mnemonic effect of simple assonance and slant rhyme, which in English are largely mid-word phenomena, would be an interesting result. In the present study none of the 104 stimulus MWIs show full rhyme. One (*fly high*) shows clipped rhyme. Another

shows slant rhyme (*quick fix*). Arguably, two more (*first term*, *hard part*) show slant rhyme in rhotic varieties of English. Thus, at least 92% of the 52 assonant stimulus MWIs show simple assonance.

SEMANTIC VARIABLES

Past research has identified several semantic variables that appreciably influence the learnability of L1 vocabulary items. Only two have attracted much attention from researchers of SLVA, namely, concreteness and imageability. Both are dimensions of perceptual—more exactly, sensori-motoric—meaning. Concreteness (vs. abstractness) is defined as the subjective concreteness of what a lexeme refers to or, put differently, "the degree to which the concept denoted by a word refers to a perceptible entity" (Brysbaert et al., 2014, p. 904). Imageability is defined as the ease with which a lexeme gives rise to a sensory mental image (Paivio et al., 1968). It appears that concreteness and imageability both appreciably enhance the learnability of single L2 words. The effects of concreteness on SLVA are especially well researched (e.g., Ding et al., 2017; Mestres-Missé et al., 2014; Pichette et al., 2012; Tonzar et al., 2009; see Peters, 2020, for a concise review). However, there is also strong evidence for the importance of imageability (e.g., de Groot, 2006; de Groot & Keijzer, 2000; Ellis & Beaton, 1993; Steinel et al., 2007). Concreteness and imageability ratings have been found to correlate with each other so strongly—*r* .83 across large samples of words (e.g., Paivio et al., 1968; Lynott & Connell, 2012)—that it is common for SLVA researchers to regard these two variables as interchangeable. For example, in various studies de Groot and colleagues used imageability ratings but wrote as if they had used concreteness ratings (for further examples see Peters, 2020; Steinel et al., 2007). A measure of either variable is relevant in relation to dual coding theory (Paivio, 1969; Paivio et al., 1968), which posits the existence of two classes of mental representations (or "codes"), one purely verbal (or propositional or symbolic) and one imagistic (or iconic or analogic). This theory predicts that concrete (or imageable) lexemes have an advantage in recall over nonconcrete (or nonimageable) lexemes because, roughly, the former are accessible using verbal and imagistic representations whereas nonconcrete-nonimageable lexemes are accessible only using verbal representations. In our study we concentrated on concreteness owing to the greater availability of published concreteness ratings of MWIs that could be used to validate the new ratings collected for the present study. Regarding concreteness, the importance of the distinction between concrete and abstract or minimally concrete lexis is recognized even by authorities who have reservations about some aspects of dual coding theory. For example, with respect to single words encountered out of context and under time pressure, Barber et al. (2013) replicated earlier findings from behavioral and neuroimaging studies that suggest that the processing of a concrete word "engages a large number of networks linked with the specific sensory-motor properties of the item" whereas an abstract word "activate[s] a number of superficial associations with other words, which cannot necessarily be integrated in a unified concept," meaning that the activation process for an abstract word is comparatively "shallow" (p. 51). In the present study, stimulus items were presented out of context; as to time pressure, although learners were not under exceptional time pressure from the researchers' point of view, the learners' impressions about this may have been different. To sum up, it seems worth bearing in mind that the encountered concrete MWIs may have activated

representations of meaning in learners' minds that were considerably more substantial than the representations of meaning activated by the abstract MWIs, all else being equal.

To estimate the concreteness of our stimulus MWIs (see the final section of Appendix 3) we collected subjective ratings of concreteness on a 1 to 5 Likert scale from 19 raters, most of whom are L1 Dutch bilinguals (e.g., colleagues or graduate students of the second author) and a few of whom are L1 English speakers, mostly retired schoolteachers. To increase the reliability of the ratings, we invited raters to rate the MWIs up to three times, with a pause between ratings. To accommodate raters willing to do this, each rater was given the list of MWIs in three different randomized orders. (These three versions of the list were the same for all raters.) Twelve of the raters rated the MWIs three times and one rated them twice. The multiple sets of ratings from any one rater were averaged to yield a single set of mean ratings for that person. Finally, a mean rating across all raters was calculated for each MWI. Besides the 104 stimulus MWIs, the raters rated 11 calibrator MWIs. The latter are MWIs for which ratings are given in the database compiled by Brysbaert et al. (2014; http://crr.ugent.be/archives/1330). The calibrators were intended first of all to serve the raters as examples of the various levels of the rating scale. For example, *big toe* and *free will* were intended to serve as examples of high and low concreteness, respectively. The 11 calibrators were placed at the top of each version of the list of to-be-rated MWIs. Eight of the 104 stimulus MWIs also happen to have "Brysbaert" ratings. Thus, to validate the new ratings, it was possible to pair 19 of the new ratings (11 calibrators + 8 others) with 19 Brysbaert ratings. The correlation between the two sets of ratings, $r = .92$, is satisfactory. A qualification is that the new ratings are, on average, 4.6% lower than the Brysbaert ratings: $Mn_{\text{Brysbaert}} = 3.26$; $Mn._{\text{New}} = 3.11$; $MnDiff = 0.153$; for which Welch's *t*-test gives, $CI_{95\%} [-0.63, 0.94]$. To assess internal reliability we used the R function "splithalf.r" in the package *multicon* (Sherman & Serfas, 2011) to calculate the mean of 5,000 split-half correlations across the ratings of all raters, with the split randomly chosen each time: *Mean split-half* $r(5,000) = .93$; *Spearman-Brown corrected reliability* $= .96$; $SD = 0.11$. In line with normal procedure (e.g., Brysbaert et al.; Warriner et al., 2013), we summarized the 19 ratings for each MWI by calculating not just the mean but also the standard deviation (SD). Our instructions to raters are given in Appendix 1. The means of the concreteness ratings of the 52 assonant and the 52 nonassonant control MWIs used in our study are, respectively, 3.11 and 3.17.

A further category of lexico-semantic variables comprises affective variables, or "dimensions" of emotional meaning. These have received little attention in SLVA research despite abundant evidence from L1 research that between-word differences in emotional meaning can have appreciable effects in L1 vocabulary acquisition (Ponari et al., 2018) and on L1 lexical processing generally (e.g., Citron, 2012), including accessibility in episodic memory (e.g., Gomes et al., 2013). The most important of these affective variables seems to be valence (Warriner et al., 2013), which is defined as the degree to which the meaning of a lexeme is pleasing or displeasing (Warriner et al., 2013) or as the degree to which the lexeme has positive, negative, or neutral emotional connotations (e.g., Ponari et al., 2018). Valence varies along a continuum that is positive at one extreme, negative at the other extreme, and neutral in the middle. *Free*, *placement*, and *vomit* are, respectively, words of extremely positive, neutral, and extremely negative valence (Warriner et al., 2013). The only study we know of that has addressed the influence of valence on the learnability of L2 vocabulary was carried out by Ayçiçeği

and Harris (2004), who observed that valenced (i.e., nonneutral) vocabulary items were comparatively well remembered in posttests of free recall and recognition. To estimate the valence of our target MWIs we used Amazon Mechanical Turk (on October 10–11, 2019) to obtain 1 to 9 Likert scale valence ratings of the 104 target MWIs from 18 raters based in the United States. Two additional sets of ratings were obtained from retired EFL teachers (native speakers of English). At the same time, we also obtained ratings for six calibrators: *guardian angel*, *nerve gas*, *musical instrument*, *identity theft*, *great grandmother*, and *drunk driver*. The instructions to raters (based on instructions used by Warriner et al., 2013) are given in Appendix 2. Sixteen of the newly rated MWIs were rated by Lindstromberg (2019), whose procedures also followed those of Warriner et al., with adaptation for MWIs. The correlation between our mean per-item ratings and those of Lindstromberg is $r = .97$. An estimate of internal reliability was calculated in the same way as for the concreteness ratings: *Mean split-half r*(5,000) = .95; *Spearman-Brown corrected reliability* = .98; *SD* = 0.08. Finally, on a 1–9 rating scale a rating of 5 indicates neutral valence. So, following Clark and Paivio (2004, p. 374), we estimated "absolute valence" by subtracting 5 from each of our 104 relevant valence ratings and removing all minus signs.

USAGE-BASED VARIABLES

Recent studies of usage-based determinants of L2 vocabulary learnability have focused overwhelmingly on learners' prior experience with vocabulary items of interest, with this experience most often being estimated by the frequency of the items in a mega-corpus such as COCA. Multitudes of results show that frequency in this sense is an influential factor in situations of natural or naturalistic vocabulary learning (e.g., Ellis, 2013). Naturalistic vocabulary learning can be defined as incidental vocabulary learning that takes place in a setting of instructed SLA. In this kind of setting one pedagogical option is to alter texts (e.g., reading texts) so that vocabulary items thought to be especially worth learning occur more often than they do in the original text. Several dozen studies have investigated the degree to which the learning of targeted vocabulary items is a function of manipulated item frequency, where manipulated item frequency is the number of times an item occurs, by design, per N running words. For example, researchers may set repetitions (i.e., manipulated frequencies) to 5, 10, and 15 repetitions per N words. Frequency in this sense is also positively associated with vocabulary learning. For example, a recent meta-analysis (Uchihara et al., 2019) of 45 effect sizes reported for 26 studies found a pedagogically significant medium-size effect: $r = .34$. It is interesting though that a facilitative effect of frequency seems to be comparatively unimportant in contexts of *intentional* vocabulary learning, both with respect to single words (De Groot & Keijzer, 2000; Ellis & Beaton, 1993) and to MWIs (Lindstromberg & Eyckmans, 2019). Probably this is because in situations of intentional learning targeted items tend to be encountered the same number of times. Nevertheless, we obtained item frequencies for the present study. As our source of MWI frequencies we chose COCA, the immense size of which is a particular advantage for measuring MWI frequencies given that many of the ones that are familiar to most native speakers occur only a small number times per million words (e.g., Moon, 1998).[3] However, we obtained frequencies of individual CWs from the 51 million word SUBTLEX-US corpus of subtitles for 8,388 films and television programs (https://www.ugent.be/pp/experimen

tele-psychologie/en/research/documents/subtlexus) because this corpus appears to reflect contemporary spoken English comparatively well (Brysbaert & New, 2009). Because both MWI frequency and total frequency of the CWs are plausible co-determinants of successful free recall and because MWI frequency and total CW frequency (or, equivalently, mean CW frequency) correlate only modestly in our data ($r = .24$), both variables were taken into account in analysis of the results from the free recall posttest that is described in a later section. However, for cued recall (as in our second posttest), the separate frequencies of CW1 and CW2 may be more important than combined CW frequencies given that CW1 and CW2 correspond, respectively, to the cue word and the response word. Because the frequencies of CW1 and CW2 each correlate strongly with total CW frequency (respectively, $r = .70$ and $.71$), we omitted total CW frequency from our statistical models for cued recall. As is commonly done, we used logged frequencies rather than raw frequencies. The relevant abbreviations are logMWI.Freq, logWordFreq, logFreqWord1, and logFreqWord2.

Finally, a partly usage-based and partly semantic variable that we took into account is the strength of the association between the CWs of a MWI. In SLVA research the best-known measure of strength of association is the mutual information (MI) score, which detects any pair of words "for which the frequency of co-occurrence is a high proportion of the overall frequency of either of the pair" (HarperCollins, 2008). Thus, a MI score gives information about the extent to which a word string is a preferred word combination within the population of speakers from which the corpus was drawn. Additionally, MI correlates positively with the degree to which a word combination has a distinct meaning or function (Ellis et al., 2008). This is relevant here because meaningfulness in this sense enhances the memorability of verbal material (Baddeley, 1999/2014, pp. 78–80). Note that although we refer to MI as a partly semantic or semisemantic variable, we might just as well call it an "indirect" semantic variable because even though MI informs about meaningfulness, formulas used to calculate it contain no term for a semantic property of any kind.[4] The MI scores used in the current study were obtained from COCA on August 22–23, 2019. The mean MI scores of the 52 assonant and the 52 nonassonant control target MWIs are, respectively, 5.31 and 4.97. Because we did not set a lower MI threshold when selecting our MWIs, the scores run from -3.86 (*take roles*) to $+12.01$ (*grand slam*), where *take roles* is the only case of negative MI. Accordingly, the proportion of unique values is high: $101/104 = .97$. These two features of the data will have enhanced *a priori* statistical power to detect an effect of MI. Note that MI scores can be misleading when MWIs occur less than three or so times in a corpus (https://corpus.byu.edu/mutualInformation.asp). However, the least frequent of our MWIs, *take roles*, has a frequency of 6. At the opposite end of the range, *long way* has a frequency of 7,833. Finally, in our data the correlation between logMWI.Freq and MI is $r = .31$.

TWO INTERLINGUAL VARIABLES

An especially well-known interlingual co-determinant of vocabulary learnability is cognateness (e.g., Otwinowska, 2015; Peters, 2020). We define it as perceived resemblance of the forms and meanings of L2 and L1 translation equivalents (e.g., English *book* and Dutch *boek*). The construct of cognateness is, however, not straightforwardly applicable to MWIs. In addition, the applicability of cognateness may be greatly reduced

when L1 and L2 belong to very different language families and cultural spheres. Consequently, researchers interested in the processing and learnability of L2 MWIs have recently begun to apply a different conception of how L2 and L1 MWIs can be similar, that of "congruency." This has been defined as the degree to which a L2 MWI has a word-for-word L1 translation equivalent (Yamashita, 2018). In a clear case of congruency each pair of counterpart L2 and L1constituent words have the same core meaning, are of the same grammatical class, and occur in the same order (cf., Wolter & Gyllstad, 2013). An example is English *new car* and Japanese *atarashii kuruma*, where *atarashii* = "new" and *kuruma* = "car." In the case of Dutch and English, which are members of the same subgroup of languages (i.e., West Germanic), it is fairly frequent for MWIs that are congruent in this sense also to have similar orthographic and/or phonetic forms even when no word has been borrowed from the other language except perhaps in the remote past. Examples of very close alignment of congruency and cognateness are English *cold/hot/warm water* and Dutch *koud/heet/warm water*. Although congruency appears to be positively associated with learnability even without cognateness (Wolter & Gyllstad, 2013; Yamashita, 2018; Yamashita & Jiang, 2010), it is plausible that presence of cognateness enhances learnability additionally. To take account of any congruency and cognateness with respect to our target MWIs we solicited 1 to 9 Likert scale ratings of "similarity" between our MWIs and Dutch translation equivalents by giving instructions that allowed our raters to interpret the word *similarity* in terms of congruency and cognateness. These raters were 44 Dutch speakers (mostly with Dutch as their L1), 33 of whom were upper intermediate learners of English and 11 of whom are Dutch–English bilingual applied linguists. Good reliability of the ratings is indicated by the strong correlation between the mean ratings of the learners and the mean ratings of the applied linguists: $r = .92$. We took this correlation as a warrant to combine the two sets of ratings into one set of 44 ratings. For these 44 ratings the mean of 5,000 split-half correlations = .95; *Spearman-Brown corrected reliability* = .98; $SD = 0.10$. The instructions to raters are given in Appendix 3, along with all 104 target MWIs and their Dutch translation equivalents.[5]

THE SELECTED VARIABLES

The lexical and semilexical variables eventually taken into account in our regression modeling are as follows: assonance, concreteness, MI, number of letters, number of phonemes, orthographic NS, phonological NS, Dutch phonological NS, logMWI.Freq, logWordFreq, logFreqWord1, logFreqWord2, similarity, and (absolute) valence.

RESEARCH QUESTIONS

The research questions for the analysis reported in the remainder of this article are as follows:

1. Following a form-focused task, which types of variable show the strongest effects in subsequent form recall tests: variables of form or variables of some other type or types?
2. Which variables show effects large enough to be of practical significance with respect to learners' ability to recall MWI forms?

3. To what extent are effects of assonance moderated by the form-focused attention-direction task and the associated manipulation of intention to remember, which we describe in the following text?

In our core model "Assonance" is the focal explanatory variable while "Focus"—the soon to be described manipulation of learners' focus on forms and intention to learn—is the prime moderator variable and "Concreteness" is a covariate of particular interest. The outcome variable is a binary test score. We expanded this core model to include further quantitative variables from among those summarized in the previous section. This was done partly to estimate the effects of these additional variables, partly to control them statistically on top of the control that may have been achieved by our screening of the pool of candidate MWIs and by randomization (as described in the following text), and partly to increase the credibility of any estimates of effects of Assonance and Focus.

METHOD

PARTICIPANTS

The participating learners were 60 undergraduate language majors studying English as one of two foreign languages within the context of their applied linguistics study program at a Flemish University in Belgium. They were all Dutch-speaking students in four intact classes. Their ages ranged between 19 and 22. Their level of proficiency in English was estimated at B2 according to the Common European Framework of Reference for Languages (CEFR), which corresponds to an IELTS score of 5–6.5.

TESTS

There was no pretest, for three reasons. First, all the CWs of the targeted MWIs are included among COCA's most frequent 5,000 lemmas, which the learners were highly likely to know. Second, as shortly to be explained, we had screened out MWIs with idiomatic meanings that we thought that an appreciable fraction of the participating learners were unlikely to know. Third, we were concerned that the learners' motivation would dwindle if we asked them to engage too often with comparatively literal MWIs made up of familiar words.

There were two posttests of recall from episodic memory. For each of the 26 MWIs that a learner had encountered during the treatment, they received a score of 1 (recalled) or 0 (not recalled). Sixty learners took the 15-minute delayed posttest of free recall, meaning that this test yielded a total of $60 \times 26 = 1,560$ binary scores. Owing to five absences, the number of scores for the 1-week delayed test of cued recall was 1,430. Results of a test of recognition, which followed the cued recall test, are not relevant here and are not discussed; however, the scores from that test are available as "Score3Recog" from the figshare data repository along with other data pertaining to the current study.

MATERIALS

The 104 target MWIs all consist of two monosyllabic content words (see the final section of Appendix 3). Candidate MWIs were drawn from two sources: Ackermann

and Chen's (2013) Academic MWI List, which consists of 2,469 two-word MWIs, and from COCA. To find suitable MWIs using COCA, we searched for the commonest immediate rightward monosyllabic noun, adjective, or manner adverb collocates of the 200 most common monosyllabic main verbs in COCA. We also used COCA to find comparatively frequent immediate leftward and rightward monosyllabic adjective, common noun, main verb, and manner adverb collocates of the 100 most frequent monosyllabic common nouns. Finally, we drew MWIs from the list of 1,000 Adj-N MWIs mentioned in a previous section. The list of candidate MWIs resulting from these searches was slightly reduced to avoid floor and ceiling effects in the planned study. For example, we screened out MWIs that, in our experience, Dutch-speaking Belgian university students of upper-intermediate English proficiency might not know and also MWIs likely to be extremely memorable on account of their emotive content (e.g., *sex toy*) or possibly quite unmemorable on account of semantic blandness. Among the MWIs in the latter category were ones including the words *type*, *kind*, *sort*, and *bit* (as in *a bit cold*); as well as the words *more* and *less*; and words for cardinal or ordinal numbers, except for *first*. To guard against wide variation in scores for individual phrases, we excluded patently ambiguous MWIs as well as ones that are technical, old-fashioned, or largely restricted to one variety of English. We also excluded MWIs that include words for objects that participants might have with them or see around them (e.g., *cup*) and words for foods, body parts, clothing, animals, and people. Alliterative phrases (e.g., *hold hands*) and full rhymes (e.g., *deep sleep*) were also excluded because they too may be comparatively memorable. Additionally, whenever COCA offered a choice between a comparatively frequent singular collocate and a similarly frequent plural collocate (e.g., *late stage/ stages*), we chose the singular form. After screening, we had more than 600 candidate MWIs of which 56 display assonance. The latter group was so small because sound-repeating MWIs are comparatively likely to be idiomatic (Gries, 2011; Lindstromberg, 2020) and were comparatively likely to be excluded for that reason.

The lists of assonant and non-sound-repeating control MWIs were cast into random order. Working down from the top of each randomized list we created four sets of 26 target MWIs (half being assonant MWIs and half being controls) by following a simple algorithm that ensured that a given CW occurs only once in a set. One of the four lists is given here:

ASSONANT: *next step*, *fun stuff*, *bright side*, *large part*, *throw stones*, *quick hits*, *sad fact*, *ride bikes*, *free speech*, *tired sigh*, *blue suit*, *sweet dreams*, *late stage*
CONTROL: *short break*, *clear view*, *push hard*, *main source*, *reach high*, *warm place*, *phone call*, *rush hour*, *sure sign*, *dance club*, *light weight*, *giant trees*, *fit well*

Note, however, that each learner saw their allotted 26 MWIs in a random order, which was the same for each participant receiving that list.

PROCEDURE

The same experimental procedure was followed in each of the four classes separately. First, the learners were informed that they were about to take part in an exercise that would be followed by a memory test. Each learner was randomly allocated one of the four lists of

26 MWIs. Learners were asked to read through their list and subvocalize the MWIs one by one. The subsequent orienting tasks were then explained:

Everyone was to write 13 of the MWIs on the back of the handout showing their allotted 26 MWIs. Learners with a list labeled "different" (i.e., half of the learners) were asked to write down the MWIs that contain *different* vowel sounds in each word whereas learners with a list labelled "same" were asked to write down the MWIs that contain the *same* vowel sound in each word. It was explained to all learners that they would *only be tested on the MWIs they were asked to write*. They were also told that they could try to memorize the MWIs of their assigned type however they liked.

After 5 minutes, the papers were collected and the learners resumed their usual class activities. Fifteen minutes later, each learner was given a blank sheet of paper and asked to write on it the 13 MWIs they had been asked to remember. When they had finished this task, they were asked to draw a line underneath their responses and then to try to add, underneath the line, the 13 MWIs they had *not* been asked to memorize. All papers were collected and the teacher moved on to unrelated matters.

A week later the cued recall test was administered. It was explained beforehand that this test related only to the MWIs presented the week before. The 55 learners who were present were given a cued recall test sheet consisting of 26 items such as "*dance* _____." The learners' task was to recall a stimulus MWI beginning with the word shown and then to write the missing second word on the test sheet. All these tests sheets were then collected.

RESULTS

DESCRIPTIVE STATISTICS

The Pearson's correlation between the scores on posttests 1 and 2, $r = .56$, is typical of what might be expected given the experimental design. (In particular, owing to the test effect, performance in a later recall test tends to be influenced by performance in an earlier recall test if the same information is tested). Table 1 shows the breakdown of scores for ±Assonance by ±Focus. By comparing the totals and the means that are diagonally opposite each other in each enclosed rectangle it can be seen that (a) in each posttest focused-on assonant MWIs were recalled better than focused-on control MWIs and (b) not-focused-on assonant MWIs were recalled better than not-focused-on control MWIs. However, it should be borne in mind that the scores within any given category in this table are not independent because some scores will have come from the same learner and some learners will have contributed more scores than other learners. Accordingly, this breakdown is not directly usable to infer statistical significance.

INFERENTIAL STATISTICAL ANALYSIS

APPROACH

For inferential analysis we used mixed-effects binary logistic regression (e.g., Baayen, 2008). To carry out the calculations we used the "glmer" function in the R (R Core Team, 2019) package *lme4*, version 1.1-21 (Bates et al., 2019). The random effects were Learner and MWI. Following Harrell (2015, pp. 68–69), we adopted an approach to stepdown

TABLE 1.   The distribution of correct test scores for the two types of MWI, with per
condition totals and, in brackets, the mean total per-learner

Test	Type of MWI	Focus… on control MWIs	Focus… on assonant MWIs
1. 15-minute delayed free recall	Control	188 [6.3]	36 [1.2]
($N_{\text{Scores}} = 1{,}560$)[a]	Assonant	68 [2.3]	234 [7.8]
2. 1-week delayed cued recallb	Control	124 [4.4]	35 [1.3]
($N_{\text{Scores}} = 1{,}430$)[b]	Assonant	54 [1.9]	147 [.4]

[a]Focus on Assonance: $n_{\text{Learners}} = 30$; Focus on *absence* of Assonance: $n = 30$.
[b]Focus on Assonance: $n_{\text{Learners}} = 27$; Focus on absence of Assonance: $n = 28$.

model simplification that mitigates disadvantages (e.g., spuriously low *p* values) of sim-
plifying a maximal model until all variables showing $p > \alpha$ are eliminated. For example,
with respect to posttest 1 we began by testing a model that included all 13 of the independent
variables in which we were interested. We then reduced this model stepwise by eliminating
variables showing $p > .50$ (not $p > .05$), stopping the model simplification process when all
remaining variables were $p < .50$. In total, six variables were excluded. In order of exclusion
they were Number of Phonemes, Valence, PhonNS, Similarity, logWordFreq, and
OrthNS.[6] Because initial complex models failed to yield any coherent output, the only
interaction that was tested was Assonance × Concreteness. For the same reason, we initially
kept the random effects portion of the model simple by including no random slopes. When
models began to converge, the random effects structure was slightly elaborated to take
account of between-learner differences in response to Assonance. The procedure with
respect to posttest 2 was similar except that for any model to converge it was necessary to
include fewer independent variables in the initial model than we had included in the starting
model for posttest 1. We therefore omitted from consideration the three variables that had
been eliminated earliest in the model simplification procedure for posttest 1, namely,
Number of Phonemes, Valence, and Similarity. Table 2 shows all the fixed effects present
in the final models for posttests 1 and 2.

RESULTS

Most notable with respect to posttest 1 is the detection of effects of Focus and MI along
with an interaction between Assonance and Concreteness (see Table 2). These results
were obtained also in posttest 2; but in that test there were also statistically significant
effects of length (i.e., number of letters) and logFreqWord1. A further detail is that our
statistical analysis did not include a term for the interaction between Focus and Asso-
nance. As previously indicated, this was to permit our earliest models to include many
different simple fixed effects and still converge. Fortunately, a reviewer reminded us of
the relevance of this interaction. We therefore retrospectively enlarged our hitherto final
models for posttests 1 and 2 by adding a term for that interaction. The statistics for that
interaction are: (Test 1) *Coefficient* $= -0.45, SE = 0.49, t = -0.93, p = .355$; (Test 2) *Coeff.*

TABLE 2. Results of the mixed-effects logistic multiple linear regression analyses (α = .05)

Test		Coefficient	*SE*	*Z*	*p*
15-minute delayed free recall	Focus	2.41	0.19	13.01	<.001***
	Assonance	2.01	0.62	3.26	.001
	Concreteness	0.47	0.14	3.43	.001
	Asson × Concr	−0.41	0.18	−2.27	.024*
	MI	0.09	0.03	2.62	.009*
	Nr Letters	−0.10	0.06	−1.55	.122
	Log MWI Freq	−0.05	0.05	−1.13	.259
	Dutch PhonNS	0.01	0.01	0.70	.482
1-week delayed cued recall	Focus	1.95	0.19	10.11	<.001***
	Assonance	2.35	0.89	2.65	.008
	Concreteness	0.79	0.19	4.16	<.001
	Asson × Concr	−0.57	0.26	−2.19	.028*
	MI	0.11	0.05	2.33	.020*
	Nr Letters	−0.28	0.11	−2.50	.012*
	logFreqWord.1	−0.19	0.09	−2.21	.027*
	OrthNS	−0.02	0.01	−1.35	.180

Notes: Freq = Frequency; *PhonNS* = Phonological neighborhood size; *Nr Letters* = Number of letters; *OrthNS* = Orthographic neighborhood size.* $p <$.05, *** $p <$.001

TABLE 3. Effect sizes of the statistically significant fixed effects, with confidence intervals (CIs)

Tests	Variable	Odds ratio ($\approx d$)	95% CI
15-minute delayed free recall	Focus	11.11 (1.33)	7.73–15.97
	MI	1.10 (0.05)	1.02–1.17
	Asson × Concr	0.66 (na)	0.46–0.95
1-week delayed cued recall	Focus	7.02 (1.08)	4.81–10.24
	MI	1.11 (0.06)	1.02–1.22
	Nr Letters	0.76 (−0.15)	0.61–0.94
	logFreqWord.1	0.83 (−0.10)	0.70–0.94
	Asson × Concr	0.57 (na)	0.34–0.95

$= −0.32$, $SE = 0.66$, $t = −0.48$, $p = .630$. Otherwise, the output for the retrospectively enlarged models remained as shown in Table 2.

Table 3 shows effect sizes for the statistically significant fixed effects (excluding the two involved in the Assonance × Concreteness interaction) in measures that may be more interpretable than logistic regression coefficients. The odds ratio (OR) for Focus indicates that focused-on MWIs are likely to be freely retrieved about 11 times more often than MWIs that are not focused on—assuming that learners, MWIs, and conditions are similar to those involved in our experiment. This large effect corresponds approximately to Cohen's $d = 1.33$. The much smaller OR for MI indicates that a 1-unit increase in MI score is associated with 6% increase in the odds of retrieval. Thus, for any two MWIs there may need to be a difference between MI scores of at least 3 or 4 for the difference to be pedagogically consequential.

The observed interaction between Assonance and Concreteness merits particular attention. Figures 1 and 2 show how predicted probabilities of retrieval of control and assonant MWIs vary across the 10th, 50th, and 90th percentiles of Concreteness. The gist of these graphical displays is as follows. First, in the "Focus off" condition (Figure 1) the predicted probabilities of retrieval (which are based on actual retrievals) are very low across the board. Second, for both types of MWI Figure 1 shows that a higher level of concreteness is associated with a greater predicted probability of retrieval (PPR). Third, this same figure shows that the gaps between the PPRs for assonant and nonassonant (control) MWIs are fairly similar at each level of concreteness although the gap is narrowest by a small amount when concreteness is high. Fourth, five of the six plots in the two figures show a distinctly higher PPR for assonant MWIs than for nonassonant MWIs. The exception (Figure 2, top right) is when Focus = On and Concreteness is high. In this case, by a tiny margin, the highest PPR is associated with *non*assonant MWIs. Finally, Figure 2 shows that the PPR of *focused-on* assonant MWIs stays near .60 at each level of Concreteness while the PPR of focused-on nonassonant MWIs rises substantially across the three ascending levels of Concreteness. We come back to this and certain other noteworthy results in the text that follows.
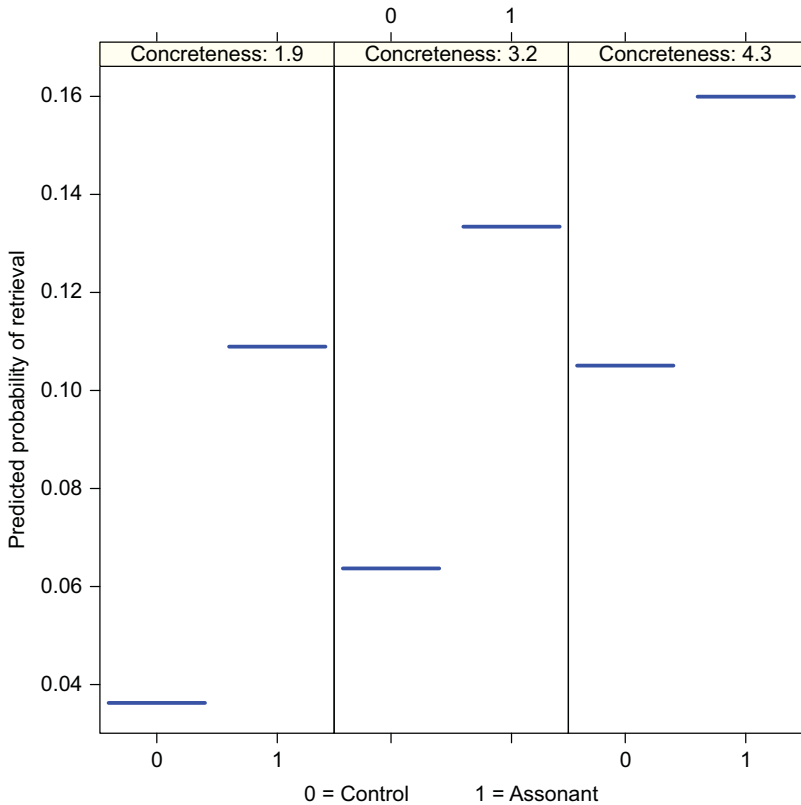

FIGURE 1.   The interaction between Assonance and Concreteness when Focus = "off" (Nscores = 780) with all additional fixed effects held at their medians. In each of the three boxes, the plot on the left relates to control MWIs and the plot on the right relates to assonant MWIs. Each horizontal bar indicates a mean predicted probability.

MODEL QUALITY

The diagnostics of model quality shown in Table 4 indicate that the final regression model for each posttest is reasonably explanatory. C-statistics were computed using the "somers2" function in the R package *Hmisc* (Harrell, 2019). A value of C above .80 indicates useful predictive ability (Harrell, 2015, p. 257). To compute values of R^2 appropriate for mixed-effects logistic regression we used the "r2" function in the R package *performance* (Lüdecke et al., 2019). As is typical for experiments of this kind, a large proportion of the variance was due to variation among the participants rather than to the fixed effects, and only a small proportion of the variation due to the fixed effects is attributable to the lexical predictors rather than to Focus (see Baayen, 2008, p. 270).

DISCUSSION

A key and possibly decisive component of the Focus = On condition was the notification that to-be-focused on MWIs would be objects of a memory test. Accordingly and as expected, this condition was associated with a sizeable positive effect on the retrieval of MWI forms. The interesting result was how general, large, and durable this effect was.
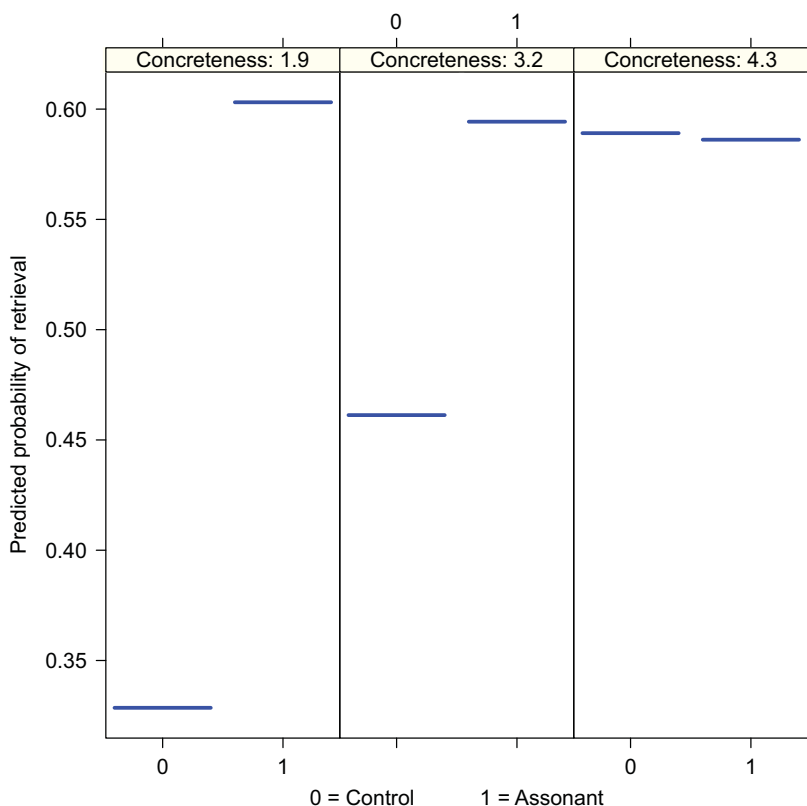

FIGURE 2.   The interaction between Concreteness and Assonance when Focus = "on" (Nscores = 780).

TABLE 4.   Indices of the quality of the two final models

Statistic	Posttest 1	Posttest 2
C statistic	.86	.88
R^2 for the fixed effects only	.30	.25
R^2 for the fixed & random effects	.44	.49

Another expected result was that learners recalled more assonant MWIs than control MWIs in both posttests and in both conditions of Focus (i.e., "on" and "off"). However, in the Focus = On condition the anticipated marked superior retrieval of the assonant MWIs was observed only with respect to MWIs of low and medium concreteness. Basically, in that condition learners' retrieval of *control* MWIs showed a marked positive association with concreteness whereas retrieval of *assonant* MWIs was fairly constant across all levels of concreteness. The reason for this difference is not clear. It may of course be a chance result arising, for example, from peculiarities of the MWIs in our study. A second possibility—given that all the CWs were likely to be familiar and that none of them seems to be strongly valenced—runs as follows. The learners who were asked to focus on nonassonant MWIs found the ones that are of low or medium concreteness to be uninteresting—the former especially so—because (a) these MWIs have comparatively insubstantial, unimageable meanings and (b) they also lack an engaging pattern of phonological similarity. Therefore—to continue the conjecture—learners who were asked to focus on the nonassonant MWIs were motivated to pay a disproportionate amount of attention to the ones whose meanings are most concrete and most imageable. In contrast, learners asked to focus on the *assonant* MWIs encountered a phonological feature of sufficient interest that these learners were motivated to engage with it regardless of the concreteness of the MWIs involved. A complementary possibility is that learners who did engage with the forms of the assonant MWIs had comparatively little processing capacity leftover to devote to processing their meanings (e.g., Barcroft, 2015), regardless of level of concreteness. If these conjectures approximate what transpired during the treatment, then the similar patterns of retrieval for the control and the assonant MWIs in the Focus = Off condition (Figure 1) can be attributed to the scant relevance of learner interest and motivation with respect to that condition. So, regarding assonance, our conclusion is that individuals who notice and pay at least a small amount of sustained attention to the presence of assonance in a MWI will benefit from doing so by remembering the form of the MWI better than would otherwise be the case—but not if the MWI is highly concrete because in that case a focus on assonance may well bring no extra mnemonic benefit. A bright side of this is that pedagogical encouragement to focus briefly on assonance (e.g., in ways described by Lindstromberg & Boers, 2008, study 3, with respect to alliteration and rhyme) seems most likely to be helpful in the case of abstract MWIs—such as our stimulus MWIs *free speech* and *prove true*—which may otherwise be comparatively unmemorable. Highly concrete MWIs such as our stimulus MWIs *ride bikes* and *throw stones* are comparatively likely to be well learned with appreciably less pedagogical attention.

 We now move on to other findings. First, our results for an interaction between Focus and Assonance are inconclusive. In particular we found no evidence that the superior

memorability of assonant MWIs compared to non-sound-repeating MWIs—conditional on concreteness—is also conditional on selective attention direction plus intention to remember: Neither assonant nor control MWIs were well remembered following exposure in the Focus = Off condition, but assonant MWIs were remembered best. Second, in posttest 2 the detected negative effects of logged cue word frequency and length in letters indicate that words that are comparatively frequent and comparatively long were less effective as recall cues than shorter, less frequent ones. A negative effect of cue word frequency may arise in part as follows. For a word to be a maximally effective prompt in a test of cued recall from episodic memory, it must be recognized. However, frequent words have a disadvantage in recognition based on episodic memory (Dewhurst et al., 2004). This disadvantage may be due to the fact that low-frequency words have comparatively distinctive forms (Pisoni et al., 1985), meaning that less frequent words tend to be more recognizable than high-frequency words. It may also matter that the forms of low-frequency words may undergo comparatively distinct encoding (Dewhurst et al.). These possibilities are relevant to MWIs because definite recognition of a cue word may increase activation of the mental representations of its collocates, especially ones encountered comparatively recently. Accordingly, our observation of a frequency effect in posttest 2 but not in posttest 1 could be due to the fact that free recall (as in posttest 1) does not involve recognition. However, a reviewer pointed out a simpler, more appealing explanation based on the concept of "cue overload" that, in memory research, is related to the well-known isolation, or von Restorff, effect: Any given high-frequency word has many more single-word collocates than a low-frequency word does, which tends to make a high-frequency word a comparatively poor cue for any specified response word.

There were a few additional variables that we expected to show small, statistically insignificant effects simply because we had taken steps to control for those variables when selecting our target MWIs. Valence is one such variable. Length is another. Nevertheless, as already mentioned, a small negative effect of length in letters was detected in posttest 2; moreover, the p value for this variable in posttest 1 is not too far above α. We had no firm expectation about a few of the other variables because little attention has been paid to them in SLVA research. Neighborhood size falls into this group, and MI too—although we have long thought it odd that mnemonic effects have been so rarely reported for MI even though it is to some extent a semantic variable. Seen as such, the durable effect of MI that we observed is unsurprising. However, it was pointed out by a reviewer that this effect might not have emerged if the stimulus MWIs had been unfamiliar to the participating learners.

Lastly, we come to Similarity (i.e., congruence and cognateness), a composite variable that we thought comparatively likely to show a positive effect. In the event, signs of an effect of Similarity were extremely weak and vague in both posttests. It may be relevant here that Casaponsa et al. (2015) found that the positive effect of cognateness declines as learners become more proficient. If this happens to be true for cognateness, it might also be true for congruence. However, the design of our study does not permit us to pursue this speculation. Another speculation is that Similarity may enhance the learnability of novel MWIs without enhancing the retrievability of the forms of familiar ones. It is likely in any case that measurement of cross-lingual similarity of MWIs is a matter calling for a good deal of further research.

SUMMARY

Our study concerned the retrievability of MWI forms subsequent to them being encountered in two conditions of Focus (i.e., selective direction of attention to features of phonological form and selective encouragement of intentional learning). The most noteworthy results are as follows. First, the study has yielded point and interval estimates of the size of the effect of Focus on subsequent retrieval of MWI forms. Second, it has provided firm evidence that assonance has an appreciable effect on the retrievability of encountered MWIs as well as evidence suggesting that the effect is conditional on the degree to which the MWIs are concrete. Interestingly, we found no evidence that the superior memorability of assonant MWIs relative to control MWIs is appreciably conditional on Focus. Additionally, where the effect of assonance exists, it seems unlikely to be substantially attributable to a task effect given that (a) learners were instructed to identify MWIs whose constituent words have different (or identical) vowels rather than identify MWIs that either do or do not assonate and (b) learners were not told how to try to remember the MWIs they had written down. Third, we detected a positive effect of Mutual Information (MI) in both posttests. Finally, an incidental result of our study is that the 104 target MWIs have been newly rated for concreteness, valence, and English–Dutch similarity. Because MWI ratings for valence are currently so rare, our new ratings for that variable may be particularly useful in validating any new MWI ratings for MWI valence that might be collected in the future.

LIMITATIONS AND CONCLUSION

One limitation of our study has already been mentioned, namely, range restriction in the case of valence and length. Another limitation is the smallness of our pool of candidate assonant targets. Were we to repeat this study, we would be less discriminating when selecting candidate target MWIs. For example, we would certainly accept disyllabic constituent words. A third limitation is that our having screened out otherwise suitable MWIs on account of apparently excessive idiomaticity means that the results of our study generalize most straightforwardly to MWIs that are comparatively nonidiomatic. Fourth, we derived our measures of valence from native speakers of English rather than from speakers of Dutch similar to the participants in our study. While this feature of our study may be nonoptimal, it is unlikely to be the sole reason why valence had so little explanatory value in our regression models. Fifth, our sample sizes (defined as $N_{\text{Learners}} \times N_{\text{MWIs}}$) could support consideration of a limited number of fixed and random effects. So, there are lexical variables that we did not attempt to take into account—for example, bigram frequency and the frequencies of orthographic and phonological neighbors (Spätgens & Schoonen, 2019). Sixth, it might be argued that the relevance of our results to L2 MWI learning in the real world is undercut by our decision not to target previously unknown MWIs. However, fairly good evidence already exists that patterns of interword phonological similarity facilitate the learning of *novel* MWIs (e.g., Boers & Lindstromberg, 2005; Eyckmans & Lindstromberg, 2017). Moreover, the ability to retrieve the forms of lexical items—which is essential for production—cannot be taken for granted even in the case of items that are familiar and well understood. It is after all well known that learners' receptive and productive abilities can be massively out of balance. More specifically, learners commonly fail to develop an awareness

of the unit status of MWIs that are made up of frequent, familiar words (Martinez & Murphy, 2011), which seems likely to retard development of an ability to fluently produce such MWIs when a need to do so arises. Another reason for our choice of familiar stimulus MWIs is that it can be counterproductive to ask learners to focus simultaneously on new forms and on the meanings of these forms (e.g., Barcroft, 2015). Finally, it might be thought that an additional limitation of our study is that in real world instructed L2 learning a teacher or materials creator who encourages learners to devote extra attention to sound repetitive MWIs reduces learners' opportunities to engage with other MWIs that may merit attention at least as much. However, *any* L2 MWI may have one or more formal or semantic characteristics with potential to make it relatively memorable but which learners may remain unaware of without a pedagogical intervention. For instance, a L2 MWI may express a vivid metaphor that can be readily clarified but that learners are unlikely to discover for themselves, or a L2 MWI may have a complete or partial L1 cognate of which learners might not be aware. The point is that when a L2 MWI is thought to be worth learning in the first place, it makes sense for a teacher or materials creator to consider exploiting *any* such affordances. Accordingly, our claim here is simply that the affordance of mnemonic interword sound repetition (e.g., assonance) should not be left out of account. Our study provides additional evidence (e.g., Boers & Lindstromberg, 2009, 2012) that time spent in alerting learners to the presence of intra-MWI sound repetition can be brief yet effective.

SUPPLEMENTARY MATERIALS

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S0272263120000315.

NOTES

[1] The list of 1,000 MWIs can be found at https://www.iris-database.org/iris/app/home/detail?id=york%3a818850&ref=search.

[2] The length data were obtained from the Clearpond database (Marian et al., 2012). The target MWIs are given in Appendix 3.

[3] MWI frequencies were collected on August 22 and 23, 2019.

[4] With respect to COCA, see the formula used to calculate MI at https://corpus.byu.edu/mutualInformation.asp

[5] Two of our target MWIs, *grand slam* and *trust fund*, have been borrowed into Dutch and have no conventional Dutch translation equivalents. Accordingly, we created two sets of Similarity scores—one set in which each of these MWIs has the maximum similarity rating and an alternate set in which each has the much lower rating corresponding to its best (but unconventional) Dutch translation equivalent.

[6] Immediately after Similarity was omitted, we tested an additional regression model that included the alternate version of Similarity referred to in note 5. This alternative version of Similarity was also omitted.

REFERENCES

Ackermann, K., & Chen, Y.-H. (2013). Developing the Academic MWI List (ACL): A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes*, *12*, 235–247. https://doi.org/10.1016/j.jeap.2013.08.002.

Ayçiçeği, A., & Harris, C. (2004). Bilinguals' recall and recognition of emotion words. *Cognition and Emotion*, *8*, 977–987. https://doi.org/10.1080/02699930341000301.

Baayen, H. (2008). *Analyzing linguistic data*. Cambridge University Press.

Baddeley, A. (1999/2014). *The essentials of human memory*. Psychology Press.

Barber, H., Otten, L., Kousta, S., & Vigliocco, G. (2013). Concreteness in word processing: ERP and behavioral effects in a lexical decision task. *Brain and Language*, *125*, 47–53. https://doi.org/10.1016/j.bandl.2013.01.005.

Barcroft, J. (2015). *Lexical input processing and vocabulary learning*. John Benjamins.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2019). *R package lme4* (Version 1.1-21). *Linear mixed-effects models using Eigen and S4.* [Computer software]. https://github.com/lme4/lme4/

Boers, F. (2020). Factors affecting the learning of multiword items. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 143–157). Routledge.

Boers, F., & Lindstromberg, S. (2005). Finding ways to make phrase-learning feasible: The mnemonic effect of alliteration. *System*, *33*, 225–238. https://doi.org/10.1016/j.system.2004.12.007.

Boers, F., & Lindstromberg, S. (2009). *Optimizing a lexical approach to instructed second language acquisition*. Palgrave-Macmillan.

Boers, F., & Lindstromberg, S. (2012). Experimental and intervention studies formulaic sequences in a second language. *Annual Review of Applied Linguistics*, *32*, 83–110. https://doi.org/10.1017/S0267190512000050.

Boers, F., Lindstromberg, S., & Eyckmans, J. (2012). Are alliterative word combinations comparatively easy to remember for adult learners? *RELC Journal*, *43*, 85–99. https://doi.org/10.1177/0033688212439997.

Boers, F., Lindstromberg, S., & Eyckmans, J. (2014a). When does assonance make lexical phrases memorable? *European Journal of Applied Linguistics and TEFL*, *3*, 93–107. https://core.ac.uk/download/pdf/55762736.pdf.

Boers, F., Lindstromberg, S., & Eyckmans, J. (2014b). Is alliteration mnemonic without awareness-raising? *Language Awareness*, *23*, 291–303. https://doi.org/10.1080/09658416.2013.774008.

Boers, F., Lindstromberg, S., & Webb, S. (2014c). Further evidence of the comparative memorability of alliterative expressions in second language learning. *RELC Journal*, *45*, 85–99. https://doi.org/10.1177/0033688214522714.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977–990. https://doi.org/10.3758/BRM.41.4.977.

Brysbaert, M., Warriner, A., & Kuperman, V. (2014). Concreteness ratings for 40,000 generally known English word lemmas. *Behavior Research Methods*, *46*, 904–911. https://doi.org/10.3758/s13428-013-0403-5.

Casaponsa, A., Antón, E., Pérez, A., & Duñabeitia, J. (2015). Foreign language comprehension achievement: Insights from the cognate facilitation effect. *Frontiers in Psychology*, *6*(588). https://doi.org/10.3389/fpsyg.2015.00588.

Citron, F. (2012). Neural correlates of written emotion word processing: A review of recent electrophysiological and hemodynamic neuroimaging studies. *Brain and Language*, *122*, 211–226. https://doi.org/10.1016/j.bandl.2011.12.007.

Clark, J., & Paivio, A. (2004). Extensions of the Paivio, Yuille, and Madigan (1968) norms. *Behavior Research Method, Instruments, Computers*, *36*, 371–383. https://doi.org/10.3758/BF03195584.

Connell, L., & Lynott, D. (2012). Strength of perceptual experience predicts word processing performance better than concreteness or imageability. *Cognition*, *125*, 452–465. https://doi.org/ 10.1016/j.cognition.2012.07.010

Davies, M. (2008). *The Corpus of Contemporary American English (COCA)*. http://www.americancorpus.org

De Groot, A. (2006). Effects of stimulus characteristics and background music on foreign language vocabulary learning and forgetting. *Language Learning*, *56*, 463–506. https://doi.org/10.1111/j.1467-9922.2006.00374.x.

De Groot, A., & Keijzer, R. (2000). What is hard to learn is easy to forget: The roles of words concreteness, cognate status, and word frequency in foreign-language vocabulary learning and forgetting. *Language Learning*, *50*, 1–56. https://doi.org/10.1111/0023-8333.00110.

Dewhurst, S., Brandt, K., & Sharp, M. (2004). Intention to learn influences the word frequency effect in recall but not in recognition memory. *Memory and Cognition*, *32*, 1316–1325. https://doi.org/10.3758/BF03206322.

Ding, J., Liu, W., & Yang, Y. (2017). The influence of concreteness of concepts on the integration of novel words into the semantic network. *Frontiers in Psychology*, *8*(2011). https://www.frontiersin.org/articles/10.3389/fpsyg.2017.02111/full.

Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *International Review of Applied Linguistics*, *47*, 157–177. https://doi.org/10.1515/iral.2009.007.

Ellis, N. (2013). *Second language acquisition*. In G. Trousdale & T. Hoffmann (Eds.), *Oxford handbook of construction grammar* (pp. 365–378). Oxford University Press.

Ellis, N., & Beaton, A. (1993). Psycholinguistic determinants of foreign language vocabulary *learning*. *Language Learning*, *43*, 559–617. https://doi.org/10.1111/j.1467-1770.1993.tb00627.x.

Ellis, N., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, *42*, 375–396. https://doi.org/10.1002/j.1545-7249.2008.tb00137.x.

Eyckmans, J., & Lindstromberg, S. (2017). The power of sound in L2 idiom learning. *Language Teaching Research*, *21*, 341–365. https://doi.org/10.1177/1362168816655831.

Forsberg, F. (2010). Using conventional sequences in L2 French. *IRAL*, *48*, 25–51. https://doi.org/10.1515/iral.2010.002.

Goldinger, S., Luce, P., Pisoni, D., & Marcario, J. (1992). Form-based priming in spoken word recognition: The roles of competition and bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 1211–1238. https://doi.org/10.1037/0278-7393.18.6.1211.

Gomes, C., Brainerd, C., & Stein, L. (2013). Effects of emotional valence and arousal on recollective and nonrecollective recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 663–677. https://doi.org/10.1037/a0028578.

Green, M. (2019). *Phonological patterns and vocabulary learning: Mora and mnemonics*. Leuven, Belgium: Paper presented at Vocab@Leuven, Department of Linguistics, KU University.

Gries, S. (2011). Phonological similarity in multi-word symbolic units. *Cognitive Linguistics*, *22*, 491–510. https://doi.org/10.1515/cogl.2011.019.

Gupta, P., Lipinski, J., & Aktunc, E. (2005). Re-examining the phonological similarity effect in serial recall: The roles of type of similarity, category cueing, and item recall. *Memory and Cognition*, *33*, 1001–1016. https://doi.org/10.3758/BF03193208

Harrell, F. (2015). *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis*. Springer.

Harrell, F. (2019). *R package Hmisc* (Version 4.2-0). [Computer software]. https://cran.r-project.org/web/packages/Hmisc/Hmisc.pdf

HarperCollins. (2008). *A guide to statistics: t-score and mutual information.* https://wordbanks.harpercollins.co.uk/other_doc/statistics.html

Laufer, B. (1997). What's in a word that makes it hard or easy? Intralexical factors affecting the difficulty of vocabulary acquisition. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary description, acquisition and pedagogy* (pp. 140–155). Cambridge University Press.

Laufer, B., & Waldman, T. (2011). Verb-noun MWIs in second language writing: A corpus analysis of learners' English. *Language Learning*, *61*, 647–672. https://doi.org/10.1111/j.1467-9922.2010.00621.x.

Li, J., & Schmitt, N. (2010). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing*, *18*, 85–102. https://doi.org/10.1016/j.jslw.2009.02.001.

Lindstromberg, S. (2020). Surplus phonological similarity in English multiword units. *Corpus Linguistics and Linguistic Theory*, *16*, 95–118. https://doi.org/10.1515/cllt-2017-0013

Lindstromberg, S. (2019). Ratings of the emotional valence and arousal of collocations and their constituent words: How can they be useful in L2 vocabulary research. *System*, *87*, 102–144. https://doi.org/10.1016/j.system.2019.102144.

Lindstromberg, S., & Boers, F. (2008). The mnemonic effect of noticing alliteration in lexical chunks. *Applied Linguistics*, *29*, 200–222. https://doi.org/10.1093/applin/amn007.

Lindstromberg, S., & Eyckmans, J. (2014). How big is the positive effect of assonance on the near-term recall of L2 collocations? *ITL International Journal of Applied Linguistics*, *165*, 19–45. https://doi.org/10.1075/itl.165.1.02lin.

Lindstromberg, S., & Eyckmans, J. (2017). The particular need for replication in the quantitative study of SLA: A case study of the mnemonic effect of assonance in collocations. *Journal of the European Second Language Association*, *1*, 126–136. https://doi.org/10.22599/jesla.26.

Lindstromberg, S., & Eyckmans, J. (2019). The effect of frequency on learners' ability to recall the forms of deliberately learned L2 multiword expressions. *ITL International Journal of Applied Linguistics*, *171*, 2–33. https://doi.org/10.1075/itl.18005.lin.

Luce, P., Goldinger, S., Auer, E., & Vitevitch, M. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception and Psychophysics*, *62*, 615–625. https://doi.org/10.3758/BF03212113.

Lüdecke, D., Makowski, D., & Waggoner, P. (2019). *R package performance* (Version 0.4.0). [Computer software]. https://easystats.github.io/performance/

Lupker, S., & Williams, B. (1989). Rhyme priming of pictures and words: A lexical activation account. *Journal of Experimental Psychology: Language, Memory and Cognition*, *15*, 1033–1104. https://doi.org/10.1037/0278-7393.15.6.1033.

Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PLoS ONE*, *7*, e43230. https://doi.org/10.1371/journal.pone.0043230.

Martinez, R., & Murphy, V. (2011). Effect of frequency and idiomaticity on second language reading comprehension. *TESOL Quarterly*, *45*, 267–290. https://onlinelibrary.wiley.com/doi/abs/10.5054/tq.2011.247708.

Mestres-Missé, A., Münte, T., & Rodriguez-Fornells, R. (2014). Mapping concrete and abstract meanings to new words using verbal contexts. *Second Language Research*, *30*, 191–223. https://doi.org/10.1177/0267658313512668.

Moon, R. (1998). *Fixed expressions and idioms in English*. Clarendon Press.

Nekrasova, T. (2009). English L1 and L2 speakers' knowledge of lexical bundles. *Language Learning*, *59*, 647–686. https://doi.org/10.1111/j.1467-9922.2009.00520.x.

Nesselhauf, N. (2003). The use of MWIs by advanced learners of English and some implications for teaching. *Applied Linguistics*, *24*, 223–242. https://doi.org/10.1093/applin/24.2.223.

Otwinowska, A. (2015). *Cognate vocabulary in language acquisition and use*. Multilingual Matters.

Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review*, *66*, 241–263. https://doi.org/10.1037/h0027272.

Paivio, A., Yuille, J., & Madigan, S. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, *76*, 1–25. https://doi.org/10.1037/h0025327.

Parkinson, D. (Ed.). (2006). *Oxford idioms dictionary for learners of English*. Oxford University Press.

Peters, E. (2020). The learning of single word items. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 125–142). Routledge.

Pichette, F., De Serres, L., & Lafontaine, M. (2012). Sentence reading and writing for second language vocabulary acquisition. *Applied Linguistics*, *33*, 66–82. https://doi.org/10.1093/applin/amr037.

Pisoni, D., Nusbaum, H., Luce, P., & Slowiaczek, L. (1985). Speech perception, word recognition, and the structure of the lexicon. *Speech Communication*, *4*, 75–95. https://doi.org/10.1016/0167-6393(85)90037-8.

Ponari, M., Norbury, C., & Vigliocco, G., (2018). Acquisition of abstract concepts is influenced by emotional valence. *Developmental Science*, *21*, e12549 https://doi.org/10.1111/desc.12549

R Core Team (2019). *R: A language and environment for statistical computing* (Version 3.6.0). [Computer software]. R Foundation for Statistical Computing. http://www.R-project.org/.

Schmitt, N. (2008). Review article: Instructed second language learning. *Language Teaching Research*, *12*, 329–363. https://doi.org/10.1177/1362168808089921.

Sherman, R., & Serfas, D. (2011). *R package multicon: Multivariate constructs* (Version 1.6). [Computer software]. https://cran.r-project.org/web/packages/multicon/index.html

Spätgens, T., & Schoonen, R. (2019). Single, but not unrelated: Key issues in researching single word items. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 511–528). Routledge.

Stamer, M., & Vitevitch, M. (2012). Phonological similarity influences word learning in adults learning Spanish as a foreign language. *Bilingualism: Language and Cognition*, *15*, 490–502. https://doi.org/10.1017/S1366728911000216.

Steinel, M., Hulstijn, J., & Steinel, W. (2007). Second language idiom learning in a paired-associate paradigm: Effects of direction of learning, direction of testing, idiom imageability, and idiom transparency. *Studies in Second Language Acquisition*, *29*, 449–484. https://doi.org/10.1017/S0272263107070271.

Storkel, H. (2004). Do children acquire dense neighborhoods? An investigation of similarity neighborhoods in lexical acquisition. *Applied Psycholinguistics*, *25*, 201–221. https://doi.org/10.1017/S0142716404001109.

Tonzar, C., Lotto, L., & Job, R. (2009). L2 vocabulary acquisition in children: Effects of learning method and cognate status. *Language Learning*, *59*, 623–646. https://doi.org/10.1111/j.1467-9922.2009.00519.x.

Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language Learning*, *69*, 559–599. https://doi.org/10.1111/lang.12343.

Warriner, A., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*, 1191–1207. https://doi.org/10.3758/s13428-012-0314-x.

Wolter, B., & Gyllstad, H. (2013). Frequency of input and L2 collocational processing. *Studies in Second Language Acquisition*, *35*, 451–482. https://doi.org/10.1017/S0272263113000107.

Yamashita, J. (2018). Possibility of semantic involvement in the L1-L2 congruency effect in the processing of L2 collocations. *Journal of Second Language Studies*, *1*, 60–78. https://doi.org/10.1075/jsls.17024.yam

Yamashita, J., & Jiang, N. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly*, *44*, 647–668. https://onlinelibrary.wiley.com/doi/abs/10.5054/tq.2010.235998.

Yap, M. J., & Balota, D. A. (2015). Visual word recognition. In A. Pollatsek & R. Treiman (Eds.), *Oxford library of psychology. The Oxford handbook of reading* (pp. 26–43). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199324576.013.4

APPENDIX 1

INSTRUCTIONS TO RATERS OF MWI CONCRETENESS

INTRODUCTION

Because some of the people we planned to recruit are laypeople, our instructions are worded colloquially. We counted on the applied linguists being understanding about this. We have omitted the introductory comment about the study and the closing thanks for participation.

THE INSTRUCTIONS

"CONCRETE" WORDS refer to things or actions that we directly experience in the real world through one of our five senses. We learn the meanings of concrete words by moving touching, picking up, holding, looking, seeing, hearing, smelling, tasting.... The easiest way to explain a concrete word is by showing.

 # To explain *chair*, you point at a chair or at a picture of a chair.

 # For the word *sweet*, give someone sugar to taste.

 # For *jump*, just jump.

"ZERO CONCRETE" WORDS are just about ideas, e.g., the word *if*.

Most words fall somewhere in between being completely concrete and being zero concrete. Let's call them MIDDLE WORDS. These we can learn to some extent through physical experience. But to learn them completely we may need a verbal definition, or we may need to hear how people around us use these words in conversation.

Here are some typical ratings for single words:

So, what is the task we're hoping you might do? It is to work down the column of two-word phrases and think about how concrete the meaning of each whole phrase is for you.

To give your rating could you key a number in opposite the phrase in the yellow column. [i.e., a highlighted column on a spreadsheet].

baby	5	← Extremely Concrete
airline	4	
appointment	3	
secrecy	2	
spirituality	1	← Near Zero Concrete

The rating scale runs from 1 to 5 with no other numbers in between … that is, no decimals or fractions, please. Incidentally, the scale starts with 1 not 0 because we doubt than any of these phrases would get a 0 rating. Estimated average time per phrase is 3–4 seconds but take as long as you like. In fact, it might be really fun and a special treat to rate the phrases two or even three times (after a bit of a delay each time). To do this just click the button at the bottom of this page which says "2nd.Rating" or "3rd.Rating."

It takes 6 or 7 minutes to rate all the phrases the first time.

APPENDIX 2

INSTRUCTIONS TO RATERS OF MWI VALENCE

INTRODUCTION

The ratings were solicited from Amazon Mechanical Turk (AMT) workers with AMT's expert qualification.

THE INSTRUCTIONS

You are invited to take part in a study of how people respond emotionally to different two-word phrases. You will use a scale to rate how you feel about the meaning of each phrase.

In total there are 110 phrases to rate. The scale ranges from 1 (totally negative) to 9 (totally positive). In the lower part of this scale the meaning of the phrase makes you feel *unhappy*, *annoyed*, *unsatisfied*, *sad*, or *despairing*. Indicate feeling COMPLETELY unhappy by typing 1. The higher part of the scale is for when the meaning makes you feel *happy*, *satisfied*, *contented*, *hopeful*. Indicate feeling completely happy by typing 9. Numbers in between 1 and 9 are for intermediate levels of feelings. If the phrase is completely neutral for you (neither happy nor sad), choose the middle of the scale, i.e., rating 5.

Please work at a rapid pace and don't spend too much time thinking about each phrase. Make your ratings based on your first and immediate reaction as you read each phrase.

Please type your single number rating in the box to the right of the phrase. No decimals, please.

This assignment normally takes less than 10 minutes to complete. Reminder: 1 = totally NEGATIVE ... 9 = totally POSITIVE ... 5 = NEUTRAL.

APPENDIX 3

INSTRUCTIONS TO RATERS OF CROSS-LINGUAL MWI SIMILARITY, WITH A LIST OF THE TARGET MWIS AND DUTCH TRANSLATION EQUIVALENTS

INTRODUCTION

The rating sheet showed a "best" Dutch translation equivalent for each of the target English MWIs. The English–Dutch translation equivalents are given just after the instructions to raters. On the rating sheets, these pairs were in random order; but they are given in the following text in the order list 1, list 2, list 3, and list 4. In each of the four lists, the first 13 English MWIs are assonant.

THE INSTRUCTIONS

Please use a 1 to 9 whole number rating scale to indicate the extent to which you think/feel that the English phrase and the corresponding Dutch phrase resemble each other.

1 = Minimum resemblance. 9 = Maximum resemblance. Naturally, intermediate numbers are for intermediate degrees of resemblance.

Example 1: *broad shoulders* = "brede schouders." If you saw each of the two English words separately, you would probably also translate them as "breed" and "schouders." The resemblance of both elements of the collocation is very strong and you could therefore award it 9.

Example 2: *go nuts* = "gek worden." If you saw each of the two English words separately, you would probably translate *go* as "gaan" and *nuts* as "noten." Neither of the words of the English phrase has a direct equivalent in the Dutch translation of the phrase. You could therefore award it "1."

large part[1]	groot deel
next step[1]	volgende stap
late stage[1]	laat stadium
throw stones[1]	stenen gooien
bright side[1]	zonnige kant
sad fact[1]	triest feit
quick hits[1]	een snel resultaat
ride bikes[1]	fietsen
fun stuff[1]	plezierige dingen
free speech[1]	vrijheid van meningsuiting
blue suit[1]	blauw pak
tired sigh[1]	vermoeide zucht
sweet dreams[1]	slaap zacht
reach high[1]	hoog reiken
push hard[1]	hard duwen
short break[1]	korte pauze
light weight[1]	lichtgewicht
sure sign[1]	zeker teken
clear view[1]	duidelijk zicht
main source[1]	hoofdbron
phone call[1]	telefoongesprek
rush hour[1]	spitsuur
giant trees[1]	gigantische bomen
warm place[1]	warme plaats
fit well[1]	past goed
dance club[1]	dansclub
main stage[2]	hoofdpodium
taste great[2]	smaakt goed
dead end[2]	doodlopende straat
wide smile[2]	brede glimlach
blow smoke[2]	rook blazen
life style[2]	levensstijl
change shape[2]	van vorm veranderen
grow old[2]	oud worden
block shots[2]	doelpogingen verhinderen
tap dance[2]	tapdansen
tax plan[2]	belastingsplan
vote no[2]	nee stemmen
gain weight[2]	aankomen, dikker worden
long way[2]	lange weg
black rock[2]	zwarte rots
print ads[2]	reclame printen
drug use[2]	druggebruik
bad luck[2]	tegenslag
lie still[2]	stil liggen
stay home[2]	thuisblijven
sound good[2]	klinkt goed
thick brush[2]	grove borstel
first thought[2]	het eerste idee
show trends[2]	een trend laten zien
knock wood[2]	afkloppen
score high[2]	hoog scoren
fly high[3]	hoog vliegen
fast track[3]	snelle baan

English	Dutch
tired smile[3]	vermoeide glimlach
light fires[3]	vuren aansteken
right side[3]	rechterzijde
get help[3]	hulp zoeken
smell fresh[3]	frisse geur
prove true[3]	aantonen
sweep clean[3]	schoonvegen
cool blue[3]	koel blauw
play games[3]	spelletjes spelen
trust fund[3]	beheerd fonds
bed rest[3]	nachtrust
pay bills[3]	rekeningen betalen
late spring[3]	late lente
wide range[3]	breed spectrum
sign books[3]	boeken signeren
core skills[3]	kernvaardigheden
buy stocks[3]	aandelen kopen
scream loud[3]	luid krijsen
full year[3]	vol jaar
main task[3]	hoofdtaak
spend hours[3]	uren doorbrengen
fun ride[3]	plezierrit
lose weight[3]	gewicht verliezen
take roles[3]	een rol opnemen
first term[4]	eerste termijn
grand slam[4]	reeks toernooien
quick fix[4]	kant-en-klare oplossing
act fast[4]	snel han+delen
sink ships[4]	schepen doen zinken
thin strips[4]	dunne strookjes
wipe dry[4]	droogvegen
real peace[4]	echte vrede
fight crime[4]	criminaliteit bestrijden
hard part[4]	hard deel
toll road[4]	tolweg
good look[4]	ferme blik
trade places[4]	van plaats wisselen
plant corn[4]	mais planten
free time[4]	vrije tijd
play roles[4]	rollen spelen
poor health[4]	slechte gezondheid
sad truth[4]	een trieste waarheid
change course[4]	van richting veranderen
switch gears[4]	schakelen
light bulb[4]	gloeilamp
old age[4]	vergevorderde leeftijd
field trip[4]	uitje
fat cells[4]	vetcellen
lean close[4]	aanleunen
take charge[4]	leiding nemen